AD-A043 180 NAVY EXPERIMENTAL DIVING UNIT PANAMA CITY FLA F/G 9/2

ANALYSIS OF MULTIVARIATE DATA: A COMPUTER PROGRAM.(U)

MAR 77 R C CARTER

UNCLASSIFIED NEDU-2-77 NL

1 OF 1

AD A043180

END

DATE FILMED

9-77

DDC

NAVY EXPERIMENTAL DIVING UNIT

7

B.S.

REPORT 2-77

ANALYSIS OF MULTIVARIATE DATA: A COMPUTER PROGRAM

Distribution Unlimited

GOLDEN ANNIVERSARY 1927-1977

EXPERIMENTAL DIVING UNIT

UNITED STATES NAVY

PANAMA CITY, FLORIDA 32407

DEPARTMENT OF THE NAVY
NAVY EXPERIMENTAL DIVING UNIT
PANAMA CITY, FLORIDA 32407

IN REPLY REFER TO:

14 NEDU-2-77

NAVY EXPERIMENTAL DIVING UNIT

REPORT 2-77

6 ANALYSIS OF MULTIVARIATE DATA: A COMPUTER PROGRAM.

By:

10 Robert C. Carter, Jr

11 25 Mar[redacted]977

12 20p.

9 Final rept.,

Approved for public release; distribution unlimited.

Submitted by:	Reviewed by:	Approved by:
ROBERT C. CARTER, Jr. LT, MSC, USN Experimental Psychologist	W. H. SPAUR CAPT, MC, USN Senior Medical Officer	C. A. BARTHOLOMEW CDR, USN Commanding Officer

253650

6

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NEDU REPORT 2-77	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ANALYSIS OF MULTIVARIATE DATA: A COMPUTER PROGRAM		5. TYPE OF REPORT & PERIOD COVERED FINAL
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) ROBERT C. CARTER, LT, MSC, USN		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Experimental Diving Unit Panama City, Florida 32407		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 25 March 1977
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multivariate Data Analysis
Computer Program

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A computer program is presented which calculates multivariate analyses of variance. The method used in the program is described, and an application of program is illustrated. The method is appropriate when several variables are measured in each of several samples, and it is hypothesized that the average of each variable is the same in all samples.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

INTRODUCTION

This report describes a computer program called MANOVA which performs multivariate analysis of variance. Multivariate analysis of variance is a technique for finding differences between samples when more than one variable is measured in each sample. This program will evaluate as many as 10 samples and 10 variables.

MANOVA deals with independent and dependent variables, so it is important to distinguish between them. Independent variables in an experiment are those which an experimenter controls; they are the treatments. Dependent variables are those which are allowed to take any values in response to the treatments. The dependent variables are the criteria by which the experimenter judges whether or not his treatments have had an effect. MANOVA is designed to be used primarily when there are multiple dependent variables. However, MANOVA also may be applied in some univariate analyses (Kirk, 1968, p. 143).

An example of an experiment in which MANOVA may be used is comparison of three scuba regulators. The independent variable might be regulator type, and the dependent variables could be maximum ergometric work produced by a diver using each regulator, and work of breathing from each regulator. Both of these dependent variables would be measured at each value of the independent variable.

After the data are gathered, the experimenter must decide whether the observed differences between regulators are due to chance or are due to real differences. MANOVA, using both of the dependent variables, will detect real differences which may not be detected by statistical tests which use only one of the dependent variables.

USER INTERFACE

PROGRAM OUTPUT

MANOVA generates several statistics including a chi-square and its degrees of freedom. These two statistics may be used to decide whether observed differences between samples are due to chance.

The number of degrees of freedeom is determined by the number of dependent and independent variables. The magnitude of chi-square depends upon the data. In general, the larger chi-square is, assuming a particular number of degrees of freedom, the smaller is the probability (p) that the observed differences between the samples are due to chance. A common criterion for decision that the difference is not due to chance is that $p < .05$. To determine whether or not this criterion is met by a particular set of data one may use MANOVA to generate a chi-square, and then consult a chi-square table, such as the one in Appendix A. If the obtained value exceeds the tabulated value then it may be assumed that a real difference is reflected in the samples. The preceding may be all the knowledge that some readers require to use MANOVA.

MANOVA also produces the mean values of the dependent variables in each treatment, and the grand mean of each dependent variable. In addition, the variances, covariances, and correlations of these variables are tabulated. If the covariance matrices of the dependent variables are not the same in all of the samples, the program will alert the user. Equality of the sample covariance matrices is an assumption of MANOVA.

PROGRAM INPUT

MANOVA is designed to interact with the user. A sample dialogue is shown in Appendix B. It is assumed that the user is working through a computer terminal, on a real-time basis. However, the program could be modified to read data from a storage device.

When MANOVA is on line, the printer will display:

"Write the number of treatment levels"

The user should respond by typing the number of treatments (or equivalently, the number of levels of the independent variables, or the number of samples). This number can be as large as 10. A carriage return should follow the number. The program will respond with:

"Write the number of variables"

The user should answer by typing the number of dependent variables, followed by a carriage return. This number, too, may be as large as 10. Next, the program will ask the user to:

"Write the number of observations in Treatment 1"

The user should indicate the number of multivariate data points (maximum 25) in the first treatment. In our regulator test example, if 20 readings were taken of work of breathing and and maximum ergometric work using the first regulator, then the user would type 20 in response to the program's query.

The program will then ask the user to enter the appropriate number of data for each variable in Treatment 1. The user should respond by typing the data, one at a time, following each by a carriage return.

When all the data for each variable have been entered, the program will ask if any data have been entered in error:

"Any errors? Y/N"

The user should type Y for yes, or N for no. If anything but N is typed, the program will guide the user through a data correction routine.

This same data entry paradigm is followed for each treatment level, until all the data has been entered. The program will begin automatically displaying its output after the last datum is typed.

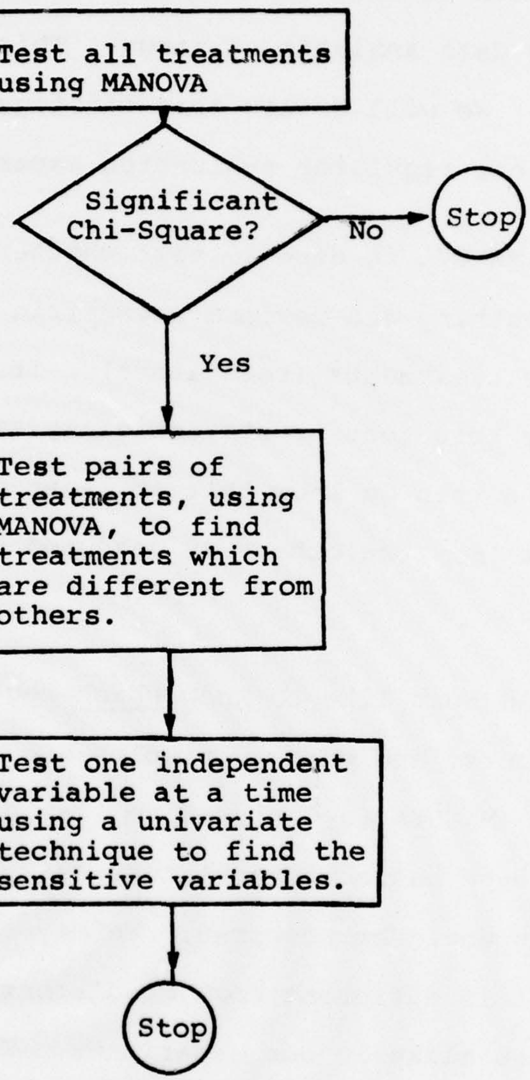
DATA ANALYSIS STRATEGY

MANOVA can be used as a tool for implementation of a multivariate data analysis strategy. This strategy is outlined in Figure 1. We will follow this strategy as it might be employed in our regulator evaluation experiment.

First, MANOVA is used to test whether the two variables (work of breathing and maximum ergometric work) are affected by the three treatments (regulators). If the chi-square generated by this test is statistically significant according to Appendix A then we know that at least one of the regulators is different from the others as measured by at least one of the variables.

We would next like to know <u>which</u> regulators are different from the others, and which variables are sensitive to that difference. MANOVA may be used to test each pair of regulators. There are three pairs of regulators (1&2, 2&3, 1&3) so three MANOVA tests would be required. We may find, for instance, that regulator #1 is different from regulators #2 and #3, but that #2 and #3 are alike in our experiment. The magnitude and direction of the differences among the regulators may be found by perusal of the mean values produced by MANOVA.

Figure 1. **DATA ANALYSIS STRATEGY**



If we are further interested in which variables are sensitive to the difference between regulators, we could perform univariate tests, such as t-tests or analyses of variance, using each variable singly. Those variables which produce significant tests are the most sensitive ones. This kind of analysis helps an experimenter to decide which variables to measure in the future. Additional insight into the relationship between the dependent variables may be obtained from study of the tables of covariances and correlations produced by MANOVA.

THEORY OF THE METHOD

MANOVA calculates a chi-square statistic. Chi-square is the distribution of the square of a variable which has a normal distribution. Hence, if $\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ is normally distributed, then $\frac{(\bar{x} - \mu)^2}{\frac{s^2}{n}}$ has a chi-square distribution with one degree of freedom. If there are k samples, then $\sum_{i=1}^{k} \frac{(\bar{x}_i - \mu)^2}{\frac{s^2}{n_i}}$ has a chi-square distribution with k degrees of freedom. If μ, the population mean, is estimated by $\bar{x}.$, the sample grand mean, then $\sum_{i=1}^{k} \frac{(\bar{x}_i - \bar{x}.)^2}{\frac{s^2}{n_i}}$ has a chi-square distribution with k-1 degrees of freedom.

In the multivariate case there are p independent variables. $\bar{x}_i$ and $\bar{x}\cdot$ (the treatment mean of the $i^{\underline{th}}$ treatment and the grand mean) are p-dimensional vectors, S is a pxp pooled covariance matrix, and n_i is the number of observations in the $i^{\underline{th}}$ sample. By analogy with the univariate case,

$$\sum_{i=1}^{k}(\bar{x}_i - \bar{x}.)\left(\frac{S}{n_i}\right)^{-1}(\bar{x}_i - \bar{x}.)$$

has a chi-square distribution with p (k-1) degrees of freedom. This method of computing a multivariate analysis of variance is equivalent to Hotelling's T^2 test (Morrison, 1967) when there are two samples, and the method is valid as a univariate test as well. It is the multivariate analysis of variance method recommended by Howe (1972).

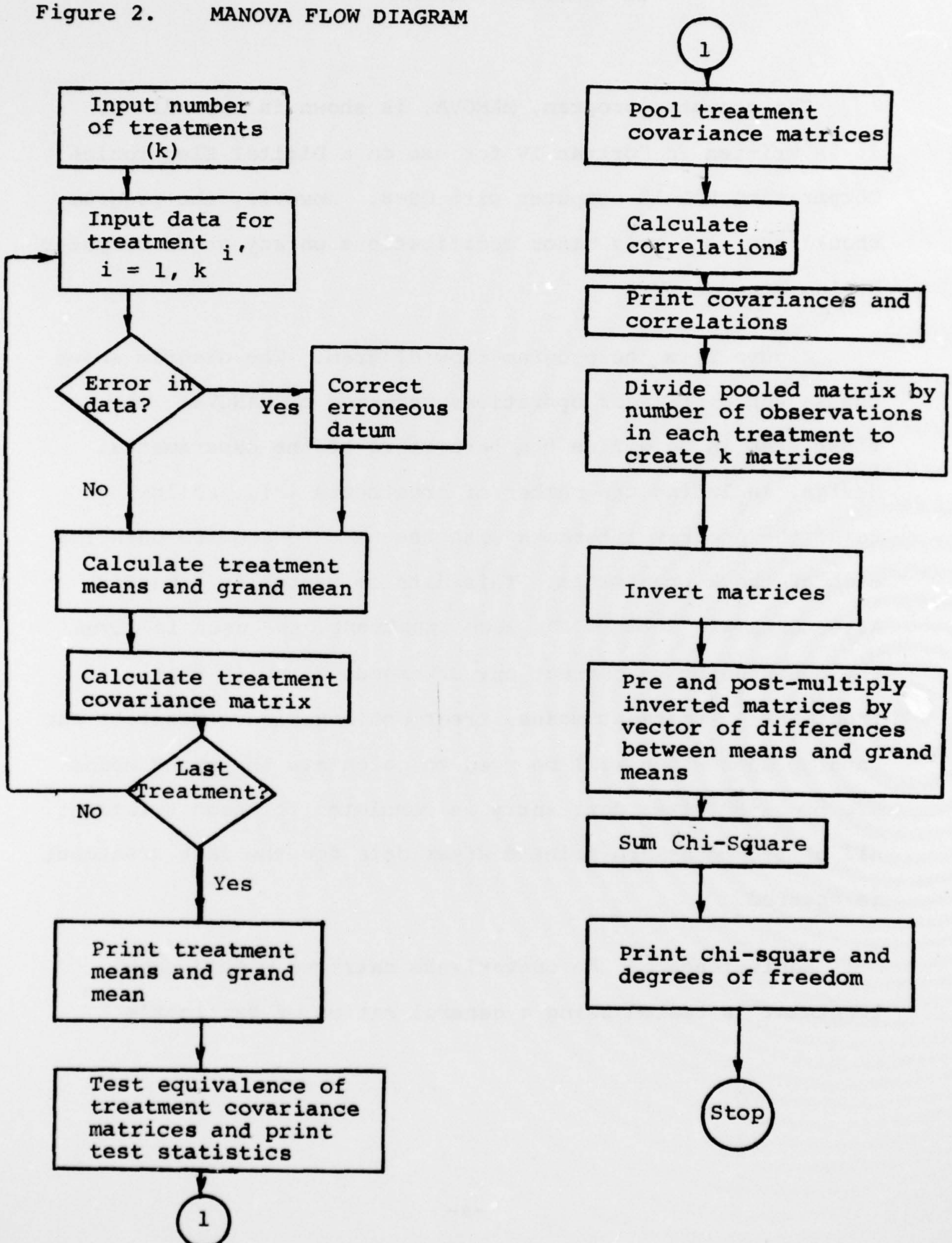
THE COMPUTER PROGRAM

The computer program, MANOVA, is shown in Appendix C. It is written in Fortran IV for use on a Digital Electronics Corporation PDP-12 computer with OS-8. However, the program should run with only minor modifications on any other computer having Fortran IV.

Figure 2 is the program flow diagram. The diagram shows the sequence of major operations peformed by MANOVA. The first step is to define the parameters of the experimental design, including the number of treatments (k). Following this, the program interacts with the user to acquire data for each of the k treatments. This data is stored in a matrix. After data are entered for each treatment, the user is given the opportunity to correct any erroneous datum in that treatment. Treatment means, treatment covariance matrices and running sums which will be used to calculate the grand means are produced after data entry is completed for each treatment. All of the means are printed after data for the last treatment is entered.

Equivalence of the convariance matrices within each treatment is tested using a generalization of Bartlett's

Figure 2. MANOVA FLOW DIAGRAM



test (Morrison, 1967, p. 152). The results of the test are printed. The covariance matrices are then combined to produce a pooled covariance matrix which is used in all subsequent calculations involving covariance. The covariance and correlation matrices are then printed.

Finally, the chi-square for differences between treatments is calculated. The pooled covariance matrix is divided by the number of observations in each treatment, and the resulting matrices are inverted using subroutine MINV (International Business Machines Corporation, IBM Scientific Subroutines, 1970). Each of these inverted matrices are pre and post-multiplied by the vector of differences between the grand means and the treatment means. These bilinear forms are added together to form the chi-square. This statistic and its number of degrees of freedom are printed.

APPENDIX A

TABLE OF MINIMUM CHI-SQUARE VALUES

($p \leq .05$)

Degrees of Freedom	1	2	3	4	5	6	7
Chi-Square	3.84	5.99	7.82	9.49	11.07	12.59	14.07

Degrees of Freedom	8	9	10	11	12	13	14
Chi-Square	15.51	16.92	18.31	19.68	21.03	22.36	23.69

Degrees of Freedom	15	16	17	18	19	20	21
Chi-Square	25.00	26.30	27.59	28.87	30.14	31.41	32.67

Degrees of Freedom	22	23	24	25	26	27	28
Chi-Square	33.92	35.17	36.42	37.65	38.89	40.11	41.34

Degrees of Freedom	29	30	40	50	60	70	90
Chi-Square	42.56	43.77	55.76	67.51	79.08	90.53	113.15

APPENDIX B

```
R FRTS
MANOVP$
WRITE THE NUMBER OF TREATMENT LEVELS
2
WRITE THE NUMBER OF VARIABLES
2
WRITE THE NUMBER OF OBSERVATIONS IN TREATMENT 1
2
TYPE  2 DATA FOR TREATMENT 1 , VARIABLE  1
2
3
ANY ERRORS? Y/N
N
TYPE  2 DATA FOR TREATMENT 1 , VARIABLE  2
2
3
ANY ERRORS? Y/N
N
WRITE THE NUMBER OF OBSERVATIONS IN TREATMENT 2
2
TYPE  2 DATA FOR TREATMENT 2 , VARIABLE  1
2
2
ANY ERRORS? Y/N
N
TYPE  2 DATA FOR TREATMENT 2 , VARIABLE  2
2
2
ANY ERRORS? Y/N
N
VARIABLE  1, TREATMENT  1: MEAN          2.50000, GRAND MEAN           2.25000
VARIABLE  1, TREATMENT  2: MEAN          2.00000, GRAND MEAN           2.25000
VARIABLE  2, TREATMENT  1: MEAN          2.50000, GRAND MEAN           2.25000
VARIABLE  2, TREATMENT  2: MEAN          2.00000, GRAND MEAN           2.25000
ROW  1   ,COLUMN   1   ,COVARIANCE      0.2500   ,CORRELATION  1.0000008
ROW  1   ,COLUMN   2   ,COVARIANCE      0.2500   ,CORRELATION  1.0000008
ROW  2   ,COLUMN   2   ,COVARIANCE      0.2500   ,CORRELATION  1.0000008
TEST FOR DIFFERENCES IN THE TREATMENT MEANS YIELDS CHI-SQUARE    1.000
WITH DEGREES OF FREEDOM   2
```

APPENDIX C

```
 L
C        THIS IS A STATISTICAL ANALYSIS PROGRAM.  IT ALLOWS THE USER TO INPUT AND
C        ANALYZE DATA WITH SEVERAL ( UP TO 10 ) VARIABLES,
C        AND ONE TREATMENT WITH SEVERAL LEVELS.  AN EXAMPLE
C        OF SUCH DATA MIGHT BE THE VARIABLES: HEART RATE,
C        RESPIRATION RATE, AND BREATHING RESISTANCE WITH A PARTICULAR
C        DIVING APPARATUS MEASURED AT SEVERAL LEVELS OF THE
C        TREATMENT, DEPTH.  THE PROGRAM WILL TEST WHETHER DEPTH
C        AFFECTS THE VALUES OF THE VARIABLES, TAKING ACCOUNT OF THE
C        INTERRELATIONSHIPS BETWEEN THE VARIABLES.   HAVE FUN!
         INTEGER P,Z,L
         DIMENSION N(10),GMEAN(10),TMEAN(10,10),COV(10,10,10)
      C  ,CCOV(100),ISCRAT(10),ISKRAT(10),CCCOV(100),
      C  VVEC(10),VEC(10),X(10,10,25),SCOV(10,10)
100      FORMAT( ' WRITE THE NUMBER OF TREATMENT LEVELS')
101      FORMAT(I1)
102      FORMAT( ' WRITE THE NUMBER OF VARIABLES')
103      FORMAT(I2)
104      FORMAT( ' WRITE THE NUMBER OF OBSERVATIONS IN TREATMENT ',I1)
105      FORMAT(I2)
106      FORMAT( ' TYPE ',I2,' DATA FOR TREATMENT ',I1,' , VARIABLE ',I2)
107      FORMAT(F10.0)
108      FORMAT( ' ANY ERRORS? Y/N')
109      FORMAT(A1)
110      FORMAT( ' TYPE THE SEQUENCE NUMBER OF THE ERRONEOUS DATUM IN
      C  THIS TREATMENT  AND',/,' THE CORRECT DATUM. SEPARATE THE
      C  SEQUENCE NUMBER AND THE CORRECTION BY A COMMA.',/,'
      C  FOR EXAMPLE: TYPING 2,7.9 WOULD MEAN',/,'
      C  REPLACE THE SECOND DATUM IN THIS TREATMENT BY 7.9.')
111      FORMAT(I2,F10.0)
112      FORMAT( ' THE TEST FOR EQUALITY OF THE TREATMENT COVARIANCE MATRICES
      C  YIELDS AN F STATISTIC ',/,'OF ',F8.3,' WITH 1 DEGREE OF FREEDOM
      C  IN THE DENOMINATOR,',/,' AND ',F4.0,' DEGREES OF FREEDOM IN
      C  THE NUMERATOR.')
113      FORMAT( ' THE TEST FOR EQUALITY OF THE TREATMENT COVARIANCE MATRICES
      C  YIELDS A',/,' CHI-SQUARE STATISTIC OF ',F8.4,' WITH '
      C  ,F4.0,' DEGREES OF FREEDOM.')
114      FORMAT( ' TEST FOR DIFFERENCES IN THE TREATMENT MEANS YIELDS
      C  CHI-SQUARE ',F8.3,/,' WITH DEGREES OF FREEDOM ',I3)
115      FORMAT( ' ROW ',I2,'    ,COLUMN  ',I2,'    ,COVARIANCE  ',F10.4,
      C  '    ,CORRELATION ',F10.7)
123      FORMAT( ' VARIABLE ',I2,', TREATMENT ',I2,': MEAN ',F16.5
      C  ,', GRAND MEAN ',F16.5)
116      FORMAT(F12.8)
C
C        INPUT DATA
C
         WRITE(4,100)
         READ(4,101) K
         WRITE(4,102)
         READ(4,103) P
         NN=0
         GMEAN(1)=0
         DO 201 J=1,K
         WRITE(4,104) J
         READ(4,105) N(J)
         NN=NN+N(J)
         DO 205 I=1,P
41       IF(J.EQ.1.AND.I.NE.P) GMEAN(I+1)=0
```

```
      WRITE(4,106) N(J),J,I
      DO 200 Z=1,N(J)
      READ(4,107) X(I,J,Z)
      GMEAN(I)=GMEAN(I)+X(I,J,Z)
200   CONTINUE
      WRITE(4,108)
C
C     CORRECT ERRONEOUS DATUM
C
      READ(4,109) KEY
      CALL CGET(KEY,1,KEY)
      IF(KEY.EQ.14) GO TO 203
      WRITE(4,110)
      READ(4,111)Z,ERR
      GMEAN(I)=GMEAN(I)-X(I,J,Z)+ERR
      X(I,J,Z)=ERR
203   IF(J.EQ.1) GO TO 218
      TMEAN(I,J)=GMEAN(I)
      DO 202 L=1,J-1
202   TMEAN(I,J)=TMEAN(I,J)-TMEAN(I,L)
      GO TO 205
218   TMEAN(I,J)=GMEAN(I)
205   CONTINUE
201   CONTINUE
      DO 219 I=1,P
      GMEAN(I)=GMEAN(I)/NN
      DO 219 J=1,K
      TMEAN(I,J)=TMEAN(I,J)/N(J)
219   WRITE(4,123) I,J,TMEAN(I,J),GMEAN(I)
C
C     CALCULATE COVARIANCE MATRIX FOR EACH TREATMENT
C
      DO 207 J=1,K
      DO 206 KAY=1,P
      COV(1,KAY,J)=0
      DO 206 KA=1,P
      IF(KA.NE.P) COV(KA+1,KAY,J)=0
      DO 206 Z=1,N(J)
206   COV(KA,KAY,J)=COV(KA,KAY,J)+(X(KA,J,Z)-TMEAN(KA,J))*
     C(X(KAY,J,Z)-TMEAN(KAY,J))
207   CONTINUE
C
C     TEST EQUALITY OF TREATMENT COVARIANCES
C
      NIN=0
      DET=0
      DO 210 L=1,K
      DO 209 J=1,P
      DO 209 I=1,P
      II=(J-1)*P+I
209   CCOV(II)=COV(I,J,L)
      NP=P
      CALL MINV(CCOV,NP,D,ISCRAT,ISKRAT)
      IF(D.EQ.0.0000) GO TO 213
      DET=DET+N(LIP)*ALOG(ABS(D))
210   NIN=NIN+1/N(LIP)
      DO 211 L=1,K
      DO 211 J=1,P
      IF(L.EQ.1) CCCOV((J-1)*P+1)=0
      DO 211 I=1,P
      II=(J-1)*P+I
      CCCOV(II)=CCCOV(II)+COV(I,J,L)*N(L)/NN
211   IF(I.NE.P.AND.L.EQ.1) CCCOV(II+1)=0
      CALL MINV(CCCOV,NP,D,ISCRAT,ISKRAT)
      EM=NN*ALOG(ABS(D))-DET
      CIN=(1-(2*P**2+3*P-1)/(6*(P+1)*(K-1)))*(NIN-1/NN)
```

```
      CHISQU=EM+CIN
      DF=(K-1)*P*(P+1)/2
      IF(NN/K.GT.20.AND.K.GT.4.AND.P.GT.4) GO TO 212
      WRITE(4,112) CHISQU,DF
      GO TO 213
212   WRITE(4,113) CHISQU,DF
C
C     CREATE POOLED ESTIMATE OF COVARIANCE MATRIX
C
213   DO 214 L=1,K
      DO 214 J=1,P
      IF(L.EQ.1) SCOV(1,J)=0
      DO 214 I=1,P
      IF(I.NE.P.AND.L.EQ.1) SCOV(I+1,J)=0
      FED=NN-K
214   SCOV(I,J)=SCOV(I,J)+COV(I,J,L)/FED
      DO 220 I=1,P
      DO 220 J=I,P
      CORREL=SCOV(I,J)/SQRT(SCOV(I,I)*SCOV(J,J))
220   WRITE(4,115) I,J,SCOV(I,J),CORREL
      CHI=0
C
C     CREATE INVERTED COVARIANCE MATRIX FOR EACH TREATMENT
C
      DO 216 L=1,K
      ZIP=N(L)
      DO 215 J=1,P
      DO 215 I=1,P
      VEC(I)=TMEAN(I,L)-GMEAN(I)
      VVEC(I)=0
      II=(J-1)*P+I
215   CCCOV(II)=SCOV(I,J)/ZIP
      CALL MINV(CCCOV,NP,D,ISCRAT,ISKRAT)
      DO 217 J=1,P
      DO 217 I=1,P
C
C     SUM CHI SQUARES
C
217   VVEC(J)=VVEC(J)+VEC(I)*CCCOV((J-1)*P+I)
      DO 216 I=1,P
      CHI=CHI +VVEC(I)*VEC(I)
216   CONTINUE
      IF=P*(K-1)
      WRITE(4,114) CHI,IF
      END

#
```

REFERENCES

Howe, W. G., Statistician, General Electric Company, Box 1072, Schenectaty, N.Y. 12301, Unpublished lecture notes, 1972.

International Business Machines Corporation, Scientific Subroutine Package (GH20-0205-4), Technical Publications Department, 112 East Post Road, White Plains, New York 10601, 1970.

Kirk, R. E., Experimental Design: Procedures for The Behavioral Sciences, Brooks Cole Publishing Company, Belmont, Calif., 1968.

Morrison, D. F., Multivariate Statistical Methods, McGraw-Hill Book Company, New York, 1967.